Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*

Celine Moorman*†, Ling V. Sun†‡§, Junbai Wang†¶, Elzo de Wit*, Wendy Talhout*, Lucas D. Ward¶, Frauke Greil*, Xiang-Jun Lu¶, Kevin P. White‡‖, Harmen J. Bussemaker¶‖, and Bas van Steensel*‖

*Department of Molecular Biology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands; ‡Department of Genetics, Yale University School of Medicine, New Haven, CT 06520; and ¶Department of Biological Sciences and Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10027

Communicated by Steven Henikoff, Fred Hutchinson Cancer Research Center, Seattle, WA, June 18, 2006 (received for review May 31, 2006)

Regulation of gene expression is a highly complex process that requires the concerted action of many proteins, including sequence-specific transcription factors, cofactors, and chromatin proteins. In higher eukaryotes, the interplay between these proteins and their interactions with the genome still is poorly understood. We systematically mapped the *in vivo* binding sites of seven transcription factors with diverse physiological functions, five cofactors, and two heterochromatin proteins at ≈1-kb resolution in a 2.9 Mb region of the *Drosophila melanogaster* genome. Surprisingly, all tested transcription factors and cofactors show strongly overlapping localization patterns, and the genome contains many "hotspots" that are targeted by all of these proteins. Several control experiments show that the strong overlap is not an artifact of the techniques used. Colocalization hotspots are 1–5 kb in size, spaced on average by ≈50 kb, and preferentially located in regions of active transcription. We provide evidence that protein–protein interactions play a role in the hotspot association of some transcription factors. Colocalization hotspots constitute a previously uncharacterized type of feature in the genome of *Drosophila*, and our results provide insights into the general targeting mechanisms of transcription regulators in a higher eukaryote.

chromatin profiling | DamID | tiling array | transcriptional regulator

Transcription factors are proteins that control the expression of specific sets of genes. They act by binding to regulatory DNA elements in the vicinity of these genes. The target specificity of transcription factors is modulated by protein–protein interactions with other factors and by the local chromatin structure. Because of the complexity of these interactions, prediction of the *in vivo* binding sites of transcription factors based on sequence alone still is unreliable (1–3). In the past few years, experimental approaches have been developed to identify the binding sites of transcription factors in living cells, on a genomewide scale (4–6).

Systematic mapping studies in yeast have indicated that each promoter is typically bound by a small set of transcription factors. The vast majority of promoters is occupied by only a few proteins, and binding of >10 proteins to a single promoter is rare (<1%) (7, 8). Although transcription factors frequently act in a combinatorial fashion (9), these results suggest that the transcription regulatory network in yeast shows a substantial degree of "division of labor," where each factor (or a small group of factors) binds a distinct set of genes.

Genomes of higher eukaryotes are much more complex than those of yeast species, raising the question of how transcription factor-binding sites are organized in such more complex genomes. Although sequence analysis of the genomes of several higher eukaryotes has indicated that transcription factor consensus motifs tend to be clustered in the genome (10–12), it is still unclear whether these clusters are the predominant targets *in vivo*, because systematic comparative studies of the genomic-binding patterns of regulatory proteins have not been reported yet. So far, genomewide-mapping efforts of transcription factor location in flies and mammals have focused on individual proteins or on small sets of functionally related factors (4–6, 13). Thus, our general knowledge of the interplay between various transcription factors and their *in vivo* binding in the genome in higher eukaryotes still is limited.

To gain understanding of the general principles that underlie regulator–genome interactions, we initiated a systematic survey of the *in vivo* genomic-binding patterns of a broad set of regulatory proteins in *Drosophila melanogaster*. We mapped the binding of these proteins in the *Drosophila* Kc cell line by using DamID technology (14, 15) combined with genomic tiling path arrays (16). Surprisingly, we found that the genomic localization patterns of unrelated factors show a strong overlap. We identified many genomic sites ("colocalization hotspots") that are targeted by most tested transcription factors and coregulators. Bioinformatics analysis and studies with mutated protein indicate that protein–protein interactions play a role in targeting of some regulator proteins to these hotspots. These results provide insights into the complex network of regulator–genome interactions in a higher eukaryote.

Results

We began our study in *Drosophila* by mapping the *in vivo* genomic-binding sites of seven transcription factors from different classes and with diverse physiological functions: Bicoid (Bcd), GAGA factor (Gaf), Jun-related antigen (Jra), Max, odd paired (opa), Ecdysone receptor (EcR) isoform B1 (EcRB1), and its heterodimerization partner Ultraspiracle (USP) (Table 1, which is published as supporting information on the PNAS web site). Except for heterodimerization of EcRB1 and USP, no physical or functional interactions have been reported between any of these proteins (www.flybase.org).

The localization patterns of these seven factors were determined in the embryonic Kc cell line, which provides a homogeneous cell population. We used the DamID technology, which has been used to map genomic binding of a variety of proteins (14, 16–20). DamID involves the *in vivo* expression of a chimeric protein consisting of a transcription factor of interest fused to *Escherichia coli* DNA adenine methyltransferase (Dam) (15). The expression levels of the Dam-fusion proteins are kept very low to avoid mistargeting due to overexpression (14, 15). DNA in the close vicinity of the natural binding sites of the transcription factor is methylated preferentially by the tethered Dam.

GENETICS

Conflict of interest statement: No conflicts declared.

Abbreviations: Bcd, Bicoid; Dam, DNA adenine methyltransferase; DBD, DNA-binding domain; EcR, Ecdysone receptor; Gaf, GAGA factor; HP, heterochromatin protein; Jra, Jun-related antigen; Pol II, RNA polymerase II; SOM, self-organizing map; USP, Ultraspiracle.

†C.M., L.V.S., and J.W. contributed equally to this work.

§Present address: Institute of Developmental Biology and Molecular Medicine, Fudan University, 220 Handan Road, Shanghai 200433, China.

‖To whom correspondence may be addressed. E-mail: kevin.white@yale.edu, hjb@genomecenter.columbia.edu, or b.v.steensel@nki.nl.

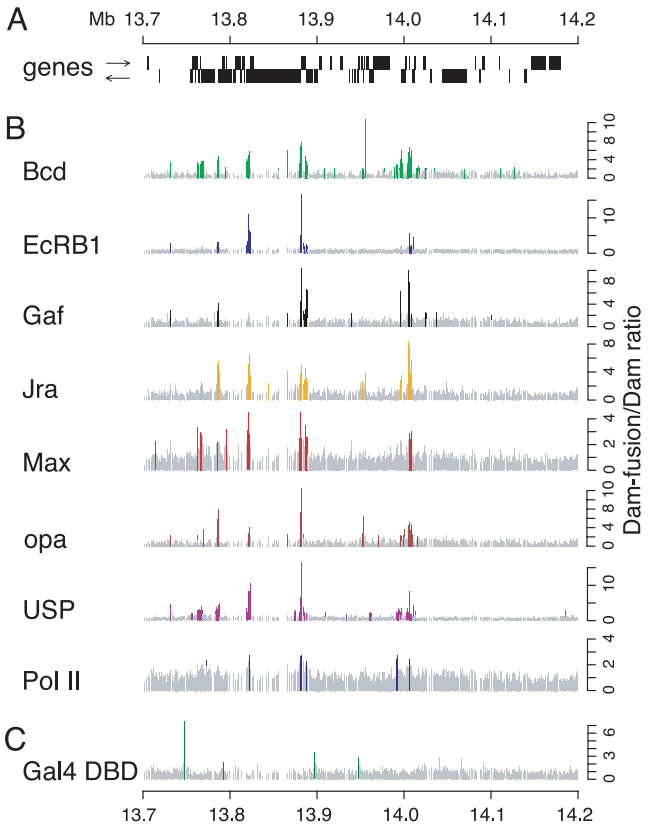

Fig. 1. Detailed view of DamID profiles generated by using genomic tiling arrays. (*A*) Predicted genes in a 0.5-Mb section of the *Adh-cactus* region (position 13,700,381 to 14,199,715 on chromosome 2L) of *Drosophila*. (*B*) DamID binding ratios of seven transcription factors and Pol II. (*C*) DamID binding ratios of Gal4p-DBD. Ratios that are statistically significant and at least 2-fold above background are highlighted by colors in *B* and *C*.

Methylated DNA fragments are isolated by using a PCR-based method, and microarrays then are used to detect the pattern of targeted adenine methylation, from which the localization pattern of the transcription factor can be deduced (14, 16, 18, 20). We used genomic tiling arrays containing 3,648 genomic fragments of 430–920 bp, together covering 2.9 Mbp of the *Adh-cactus* region of chromosome 2 of *Drosophila* (16).

The results for all seven transcription factors are displayed graphically in a detailed map of the *Adh-cactus* region (Figs. 1 *A* and *B* and 2*A*; see also Fig. 7, which is published as supporting information on the PNAS web site). The fraction of probed regions targeted by individual proteins ranged between 3–8%. Surprisingly, we found that there is a high degree of overlap between the localization profiles of these seven factors. Each factor shares at least 67% of its binding sites with one or more other factors, and many sites are targeted by all seven factors (see below). Correlation analysis showed that the similarity of the profiles is in all cases statistically highly significant (Fig. 8, which is published as supporting information on the PNAS web site). The high degree of overlap is not restricted to the 2.9 Mbp *Adh-cactus* region; we also observed it for three of the transcription factors (Bcd, Gaf, and Jra) on arrays containing ≈12,000 cDNA fragments, corresponding to ≈60% of the *Drosophila* genes (Fig. 9, which is published as supporting information on the PNAS web site). MatrixREDUCE (21, 22) analysis of the DamID profiles of Bcd, EcRB1, Gaf, Max, and Jra shows that they correlate strongly with the *in vitro* sequence specificity predicted by using their respective known binding motifs (Table 2, which is published as supporting information on the PNAS web site; also see *Materials and Methods*). USP correlates weakly with both its own binding motif and the binding motif of its heterodimerization partner, EcR (P = 0.057 and 0.035, respectively.). Using a reporter gene containing an ecdysone response element, we could also show that Dam does not interfere with the function of EcRB1 (L.V.S. and K.P.W., unpublished results). These results suggest that the Dam-fusion proteins are functional. Taken together, our results reveal a high degree of colocalization for a wide range of transcription factors that (with the exception of EcRB1 and USP) were thought to be functionally unrelated.

To further verify our DamID results, we performed ChIP, a method for detecting protein–DNA interactions that is fundamentally different from DamID (4, 5, 23). The ChIP profile of endogenous Gaf (i.e., in untransfected cells) is in good agreement with the Gaf DamID profile (Fig. 3). This agreement also was reported recently in an independent comparison (24). We also confirmed the DamID microarray data for a subset of transcription factors and genomic locations by using quantitative PCR (Fig. 10, which is published as supporting information on the PNAS web site), ruling out that our results are caused by a microarray artifact.

To investigate whether the high degree of colocalization is caused by nonspecific protein–DNA interactions, we determined the DamID profile of the DNA-binding domain (DBD) of the sequence-specific yeast transcription factor Gal4p (Fig. 1*C*). We found that Gal4p DBD binds significantly to 16 fragments on the array (corresponding to 11 loci), but there is no overlap with the localization profile of the other proteins (Figs. 1 and 8). The observed DamID profile of Gal4 DBD correlates strongly with the Gal4p consensus motif (Table 2), which confirms that the Gal4 DBD-Dam fusion protein is functional. We conclude that the colocalization of the seven *Drosophila* transcription factors is not simply due to nonspecific DNA binding.

Transcription factors rely on interactions with other proteins, such as corepressors, coactivators, and chromatin proteins, to regulate gene expression. We analyzed whether these proteins also associate with transcription factor-binding sites. Indeed, the corepressors Groucho (Gro), Rpd3, and Sin3 and the coactivator Brahma (Brm) colocalize strongly with the transcription factors and with each other (Figs. 2, 7, and 8). The heterochromatin proteins HP1a and Su(var)3-9 do not associate with the same loci as the transcription factors (Figs. 2, 7, and 8), but they do target transposable elements and a few other loci, consistent with previously reported data (16, 18). Heterochromatin protein 1c (HP1c) a euchromatic homologue of HP1a of unknown function (18, 25), shows a similar profile as the cofactors (Figs. 2, 7, and 8), suggesting that HP1c may act as a transcriptional regulator.

Visual inspection of the DamID patterns indicated that the *Adh-cactus* region contains many sites where all transcription factors and cofactors associate together. To identify such sites in a systematic and unbiased way, we applied machine learning methods to classify protein composition along the *Adh-cactus* region of *Drosophila* into a number of distinct "chromatin types." We first used a self-organizing map (SOM) algorithm (26) to generate a 2D representation of the DamID data for all 14 proteins, in which different areas correspond to genomic regions with different protein compositions (Fig. 4*A*). Next, we used adaptive K means clustering (29) to assign each probed genomic fragment to one of eight distinct chromatin types (Fig. 4*B*). This number of chromatin types was determined in an unsupervised manner (29). Although this classification is likely to represent an oversimplification of chromatin diversity, it provides a useful framework for further analysis. Chromatin types that can be distinguished are characterized by, for instance, binding by HP1a and Su(var)3-9 only ("heterochromatin"; fragments of type 5) and no binding of any protein (fragments of type 2). Fragments of type 1 show high binding by all proteins except HP1a and

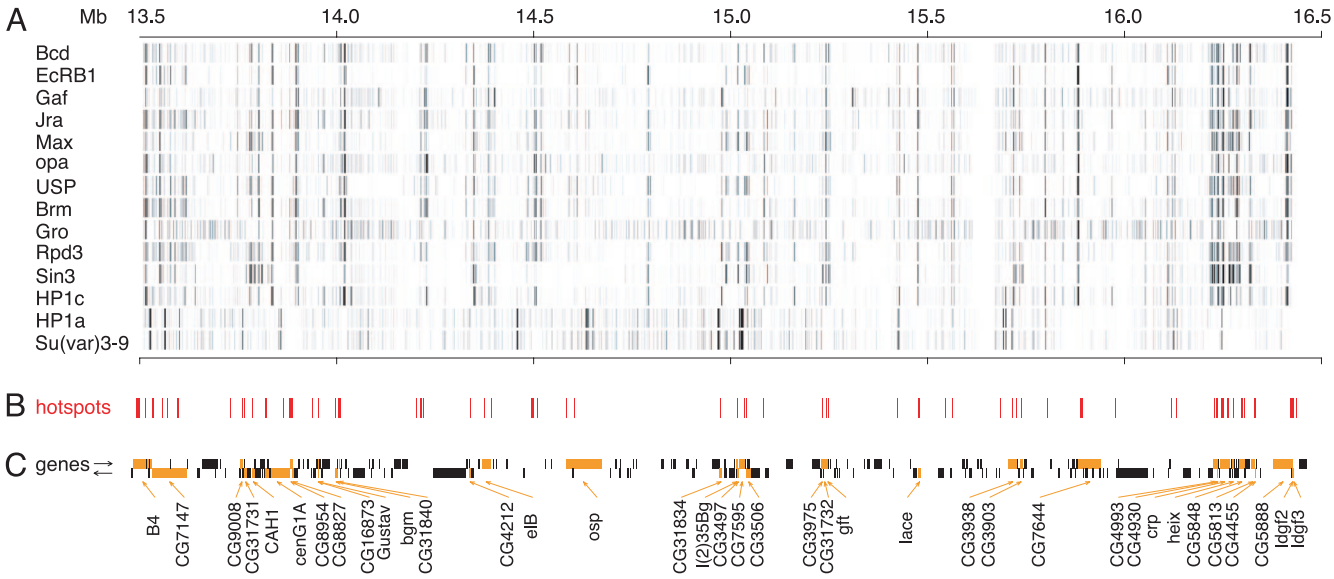

Fig. 2. Complete view of DamID profiles of the entire 2.9-Mb *Adh-cactus* region. (*A*) DamID data for all 14 *Drosophila* proteins that were studied. Each probe on the array is represented by a vertical line, the color of which indicates the relative binding of the protein (ranging from white equals background level to black equals strongest binding). (*B*) Location of hotspots, i.e., probes that are part of cluster 1 in Fig. 4*B*. (*C*) Location of all predicted genes in the probed region. Genes that overlap with at least one hotspot are highlighted in orange, and their names are indicated.

Su(var)3-9 and include 166 fragments congregating in 61 loci. We will refer to loci of type 1 as "colocalization hotspots" (Figs. 2*B* and 7). We also visualized the DamID data of GAL4-DBD and the ChIP data of Gaf in the SOM plane (Fig. 4*C*). This visualization shows clearly that GAL4-DBD as detected by DamID is completely absent from hotspots, whereas binding of Gaf as detected by ChIP is strongly enriched in hotspots, in agreement with the Gaf DamID results.

We computationally screened a collection of transcription factors of known sequence specificity (30, 31) for preferential binding to colocalization hotspots (see *Materials and Methods*). Fig. 5 shows that the hotspot regions on average have higher predicted affinity for three of the proteins that we mapped in this study, Gaf, Jra, and Max. In addition, we found motifs of several other proteins to be enriched in the colocalization hotspots (Table 3, which is published as supporting information on the PNAS web site), predicting that these proteins also bind to hotspots. Surprisingly, the motifs for Bcd, EcR, and USP are not enriched in hotspots (Fig. 5). This result suggests that these latter factors are not recruited to hotspots via protein–DNA interactions mediated by their own DNA-binding domain but rather via protein–protein interactions with one or more other DNA-bound proteins.

We investigated the mechanism of regulator targeting to hotspots more directly for the transcription factor Bcd. For this purpose, we generated DamID profiles of Bcd mutants that either lacked a functional DBD (Bcd^{K50A}; refs. 32 and 33) or consisted of the DBD alone (Fig. 7). Strikingly, Bcd^{K50A} localizes to the hotspots, whereas the DBD alone is not detected in most hotspots (Fig. 4*D*). The Bcd DBD binds significantly to 69 fragments on the array, only ≈25% of which also are bound by WT Bcd. Its DamID signal correlates significantly better with the predicted affinity for Bcd than the DamID profile of WT Bcd; by contrast and as expected, the Bcd^{K50A} DamID signal entirely lacks correlation with the Bcd consensus (Table 2). Taken together, these results suggest that protein–protein interactions help target Bcd to the hotspots, possibly in competition with the direct recruitment of Bcd to its DNA-binding sites in the genome.

We noticed that colocalization hotspots are often located in predicted genes: 41 of the 61 hotspot regions overlap with 37 genes (Fig. 2*C*). To test whether hotspots occur preferentially in actively transcribed regions, we first mapped the binding of the

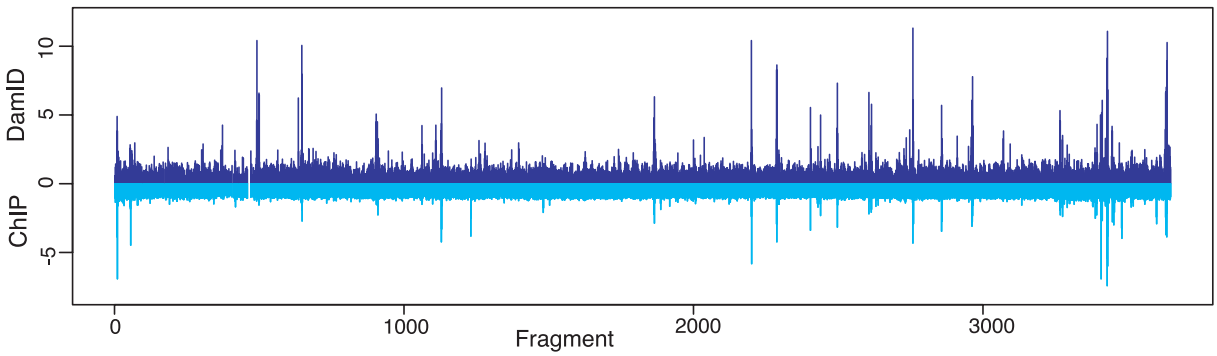

Fig. 3. Comparison of Gaf DamID to Gaf ChIP. Gaf DamID-binding ratios across the entire 2.9 Mb *Adh-cactus* region are plotted on the positive *y* axis (dark blue), and Gaf ChIP-binding ratios are plotted on the negative *y* axis (light blue).

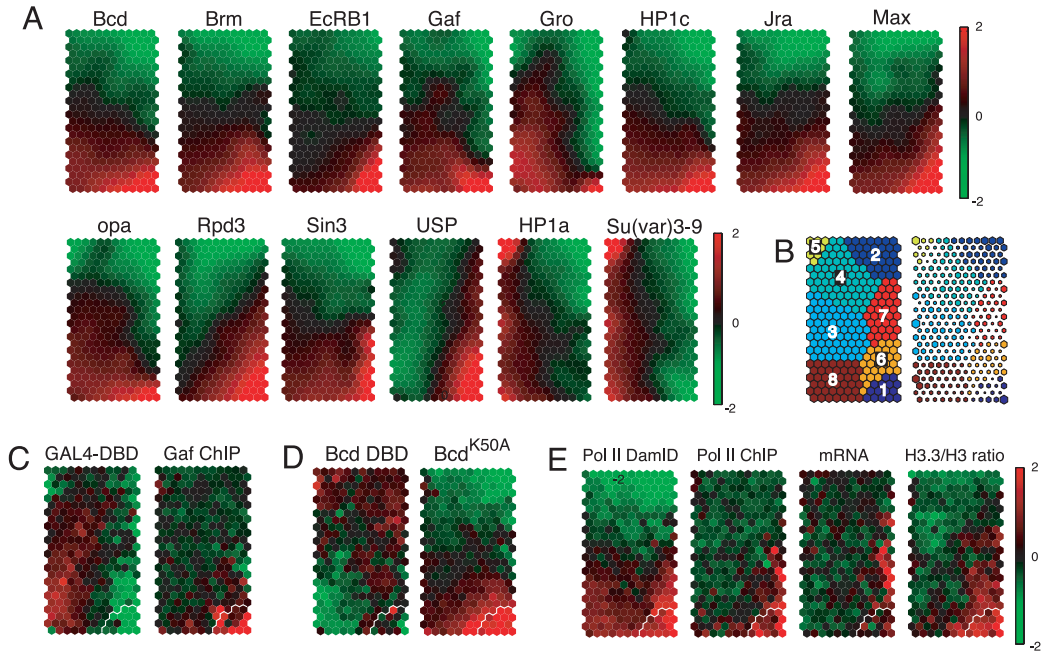

Fig. 4. Definition of hotspot regions and other types of chromatin by SOM analysis and K means clustering of DamID profiles of 14 proteins. (*A*) SOM analysis visualized for each protein in SOM component planes consisting of a matrix of 24 × 12 hexagonal nodes. Each node includes tiling array fragments with similar binding behavior across all 14 DamID profiles. Nodes that are close to each other in the SOM component plane are more alike than nodes that are farther away. The color scale represents the mean binding *Z* score in each node. (*B Left*) Further simplification of SOM data by adaptive K means clustering, resulting in the classification of the combined protein-binding patterns into eight types of chromatin, each marked in a different color. (*B Right*) The area of each node reflects the number of tiling array fragments included in that node. (*C*) DamID profile of GAL4-DBD and ChIP profile of Gaf visualized in the same SOM matrix as in *A* and *B*. (*D*) DamID profiles of Bcd DBD and Bcd^{K50A} visualized in the same SOM matrix as in *A* and *B*. Bcd^{K50A} binds strongly to the hotspots (''type 1,'' outlined by a white line), whereas Bcd DBD does not. (*E*) SOM matrix visualization of (from left to right) DamID profile of the 18-kDa subunit of Pol II, ChIP profile of the CTD subunit of Pol II (data from ref. 27), mRNA expression levels (data from ref. 27), and histone H3.3 over H3 ratios (data from ref. 28). Note the enrichment of all of these transcription-related marks in the hotspot cluster (outlined by a white line).

18-kDa subunit of RNA polymerase II (Pol II) in the *Adh-cactus* region. The DamID profile of Pol II shows strong overlap with the binding profiles of the transcription factors (Figs. 1, 7, and 8), and hotspots were significantly enriched in Pol II (Fig. 4*E*). Essentially the same result was obtained when we analyzed a previously reported Pol II-binding profile in Kc cells that was obtained by ChIP (27), again indicating that the overlap is not caused by a bias in the DamID method (Figs 4*E* and 8). Furthermore, comparison of our data to chromosomewide expression profiling in Kc cells (27) also shows that mRNA levels correlate with the binding of the transcription factors (Fig. 8) and are significantly elevated in colocalization hotspots (Fig. 4*E*).

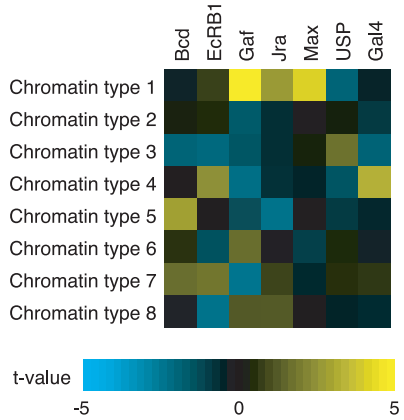

Fig. 5. Sequence signals enriched in the eight different chromatin types shown in Fig. 4*B*. Information from the TRANSFAC (30) and JASPAR (31) databases was used to construct a position-specific affinity matrix (31) representing the sequence specificity of each transcription factor. Each position-specific affinity matrix was used to predict a profile along the entire *Adh-cactus* region of the binding affinity relative to the optimal binding sequence. Difference between average for probed regions of a given chromatin type (see Fig. 4*B*) and all other regions, represented as a *t* value computed by using a paired *t* test. Note the increased predicted affinity of colocalization hotspots (type 1 regions) for Gaf, Jra, and Max.

Recent studies in *Drosophila* cells have shown that the histone variant H3.3 is specifically deposited in transcribed genes and their flanking regions (28). Analysis of these data revealed that colocalization hotspots are strongly enriched in H3.3 (Fig. 4*E*). We found that 95% of the hotspot probes have H3.3/H3 ratios that are above the mean of the entire *Adh-cactus* region ($P < 2e-16$; Student *t* test), indicating that colocalization hotspots virtually are restricted to chromatin regions containing H3.3. Taken together, our analyses indicate that hotspots occur preferentially in regions of active transcription.

We wondered whether colocalization hotspots, which we identified in Kc cells, also may be linked to sites of transcription during fly development. We therefore analyzed the developmental expression profiles of genes overlapping with hotspots by using published genomewide expression profiles from six different developmental stages (34). Strikingly, this analysis (Fig. 6) showed that during embryogenesis, hotspot-containing genes are expressed at significantly higher levels than hotspot-free genes. This difference disappears later in development. A modest increase of hotspot-associated gene expression also is observed in adult females, which may be due to the fact that female ovaries contain large amounts of oocytes with maternally loaded mRNA. These results indicate that colocalization hotspots occur preferentially in regions with increased transcription activity during early development.

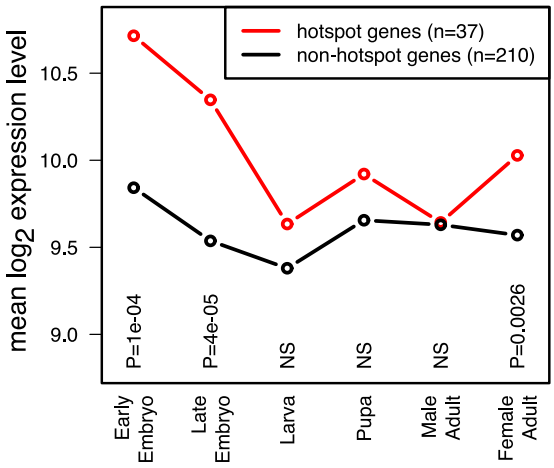

Fig. 6. Different developmental expression profiles of hotspot-containing genes (red) compared with hotspot-free genes (black) in the *Adh-cactus* region. Values indicate mean $\log_2$ expression levels (arbitrary units; data taken from ref. 34). *P* values indicate the statistical significance of the difference for each developmental stage (Wilcoxon test, Bonferroni-corrected for multiple testing); NS, not significant ($P > 0.01$).

Discussion

We report here that the target loci of several unrelated transcription factors and cofactors strongly overlap with each other in the genome of *Drosophila*. Colocalization hotspots, which are targeted by all seven transcription factors and six of eight other proteins tested, represent extreme cases of this phenomenon. Because *Drosophila* contains ≈700 transcription factors (35), we predict that many more factors will show strong colocalization and that hotspots may recruit up to a few hundred different proteins. This observation contrasts with results in yeast where overlap between transcription factors was relatively rare and hotspots were not apparent (7, 8).

The mechanism by which such a large number of proteins are recruited to specific genomic loci remains to be elucidated. Our observation that Bcd does not require its DBD for hotspot association, together with the lack of enrichment of the consensus motifs for USP, Jra, and Bcd in hotspots, indicates that at least some transcription factors do not directly bind the DNA in hotspots, but instead are recruited through protein–protein interactions. We note that the DamID maps represent the average protein occupancy over time and over many cells, and, therefore, it is quite possible that the composition of hotspot-associated complexes is variable and dynamic (36).

The function of hotspots is unclear, but at least three models can be envisioned that are not mutually exclusive. First, hotspots may represent general "sinks" or "buffers" that sequester many regulator molecules. Thus, hotspots may regulate the effective concentration of free transcription factor molecules in the nucleus. This model is in agreement with our observation that wild-type Bcd fails to bind to several loci containing the Bcd consensus motif, whereas the Bcd DBD (that is not tethered to hotspots) does bind to these loci. Second, hotspots may be directly involved in the regulation of nearby genes, similar to enhancers. Unlike classical enhancers, however, hotspots may associate with hundreds of proteins, many of which may have only a minor contribution to the overall effect on gene expression. Third, hotspots could mediate physical interactions between distant loci in the genome, as has been reported for several transcription factor-rich loci (37, 38). Regardless of what could be the regulatory function of hotspots, our results have important implications for our understanding of the mechanisms that determine the target specificity of transcription factors and for the interpretation of genomewide protein occupancy data in higher eukaryotes.

Materials and Methods

DamID and ChIP. Detailed descriptions of the methods for mapping protein location are described in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

Data Analysis. All measured ratios in the DamID and ChIP experiments were $\log_2$-transformed and normalized to the median value of all *Adh-cactus* fragments. Release 3.1 of the *Drosophila* genome sequence and annotation was used (39). For the SOM analysis, an average over replicates was calculated, and probes with one or more missing values were excluded, leaving 3,140 probes. For each protein, the data were then normalized to zero mean and unit variance (Z scores). Training of the SOM was performed as described in ref. 27 by using a hexagonal topology of 24×12 nodes. Nodes in the converged SOM then were clustered into eight sets by using a K means algorithm, the optimal number of clusters being determined by the Davies-Bouldin index (29).

To define the number of hotspots in the *Adh-cactus* region, we determined the number of clusters of contiguous fragments ("contigs") of the hotspot type (i.e., type 1, as defined by SOM analysis). If hotspot fragments were interrupted by 1–2 fragments that had missing values for all (or most) of the proteins, we regarded those hotspot fragments as part of the same contig. In this way, we found 61 contigs varying from 1 to 10 fragments.

To uncover sequence signals enriched in the loci targeted by each Dam fusion protein, we predicted the affinity of each probed region for a large number of transcription factors. To this end, we downloaded position-specific nucleotide count matrices from TRANSFAC (30) and JASPAR (31), added a pseudocount of one at each position, and normalized each position by dividing by the count for the most frequent base. The resulting position-specific affinity matrix was used to predict the (relative) affinity of each probed region for the corresponding transcription factor as described in refs. 21 and 22. MatrixREDUCE software (21, 22) was used to compute the *t* value for the regression coefficient between DamID signal and predicted affinity. The data were preprocessed by taking the residuals from a multivariate model fit based on counts for all mono- and dinucleotides, as well as the motif GATC, the site methylated by Dam. To uncover sequence signals enriched in the probed regions making up a given chromatin type, we compared the distribution of their predicted affinity with that of all other probed regions by using a paired *t* test.

To compare the Pol II ChIP and chromosomewide expression profiling data (27) to the data generated in this study, we resampled the 1.5-kb tiling fragment data from MacAlpine *et al.* (27) as follows. For each *Adh-cactus* fragment, we determined the overlapping fragment(s) on the microarray of the MacAlpine study. If there was one overlapping fragment, we assigned the mean value of that fragment to our *Adh-cactus* fragment. If two or more partially overlapping fragments were found (which was the maximum), we assigned a weighted average of the mean values of these two fragments to our *Adh-cactus* fragment. For analysis of the enrichment of H3.3, we subtracted for each 100-bp tiling array fragment in the data set from Mito *et al.* (28) the mean H3 log ratio from the mean H3.3 log ratio. The resulting H3-corrected log ratios then were resampled to match our *Adh-cactus* tiling fragments.

For the analysis of the developmental expression patterns, we used the genomewide data from Stolc *et al.* (34). Hybridization intensities of each exon probe were $\log_2$-transformed, data from replicate experiments were averaged, and exon probes corresponding to the same gene subsequently were averaged to obtain a single $\log_2$ expression level for each gene at each developmental stage. Next, these values were normalized to the mean $\log_2$

GENETICS

value, first by column (developmental stage) then by row (gene). The developmental expression levels of genes overlapping with one or more hotspots were compared with those of genes not overlapping with hotspots.

Supporting Information. Additional data can be found at Fig. 11 and Data Set 1, which are published as supporting information on the PNAS web site.

We thank Susan Parkhurst and Amir Orian (Fred Hutchinson Cancer Research Center, Seattle, WA) and David Wassarman (University of Wisconsin, Madison, WI) for plasmids; Nicolas Nègre and Giacomo Cavalli (Institute of Human Genetics, Centre National de la Recherche Scientifique, Montpellier, France) for ChIP protocol and Gaf antibody; Marcel van Batenburg and Barrett Foat for technical advice; and Melissa Davis for preparing microarrays. Some materials were received through the Drosophila Genomics Resource Center. This work was supported by a Human Frontier Science Program Grant (to B.v.S., K.P.W., and H.J.B.); National Institutes of Health Grants HG003008, CA12152, and R24 GM074105 (to H.J.B.) and R01 HG03231 (to K.P.W); the W.M. Keck Foundation (K.P.W.), the Beckman Foundation (K.P.W.), and a European Young Investigators Award (to B.v.S.).

1. Bulyk, M. L. (2003) *Genome Biol.* **5,** 201.
2. Chua, G., Robinson, M. D., Morris, Q. & Hughes, T. R. (2004) *Curr. Opin. Microbiol.* **7,** 638–646.
3. Siggia, E. D. (2005) *Curr. Opin. Genet. Dev.* **15,** 214–221.
4. Bertone, P., Gerstein, M. & Snyder, M. (2005) *Chromosome Res.* **13,** 259–274.
5. Blais, A. & Dynlacht, B. D. (2005) *Genes Dev.* **19,** 1499–1511.
6. van Steensel, B. (2005) *Nat. Genet* **37,** Suppl., S18–S24.
7. Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., Hannett, N. M., Tagne, J. B., Reynolds, D. B., Yoo, J., *et al.* (2004) *Nature* **431,** 99–104.
8. Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., *et al.* (2002) *Science* **298,** 799–804.
9. Beer, M. A. & Tavazoie, S. (2004) *Cell* **117,** 185–204.
10. Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. & Eisen, M. B. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 757–762.
11. Markstein, M., Markstein, P., Markstein, V. & Levine, M. S. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 763–768.
12. Rajewsky, N., Vergassola, M., Gaul, U. & Siggia, E. D. (2002) *BMC Bioinformatics* **3,** 30.
13. Orian, A. (2006) *Curr. Opin. Genet. Dev.* **16,** 157–164.
14. van Steensel, B., Delrow, J. & Henikoff, S. (2001) *Nat. Genet.* **27,** 304–308.
15. van Steensel, B. & Henikoff, S. (2000) *Nat. Biotechnol.* **18,** 424–428.
16. Sun, L. V., Chen, L., Greil, F., Negre, N., Li, T.-R., Cavalli, G., Zhao, H., van Steensel, B. & White, K. P. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 9428–9433.
17. Bianchi-Frias, D., Orian, A., Delrow, J. J., Vazquez, J., Rosales-Nieves, A. E. & Parkhurst, S. M. (2004) *PLoS Biol.* **2,** E178.
18. Greil, F., van der Kraan, I., Delrow, J., Smothers, J. F., de Wit, E., Bussemaker, H. J., van Driel, R., Henikoff, S. & van Steensel, B. (2003) *Genes Dev.* **17,** 2825–2838.
19. Orian, A., van Steensel, B., Delrow, J., Bussemaker, H. J., Li, L., Sawado, T., Williams, E., Loo, L. W., Cowley, S. M., Yost, C., *et al.* (2003) *Genes Dev.* **17,** 1101–1114.
20. van Steensel, B., Delrow, J. & Bussemaker, H. J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2580–2585.
21. Foat, B. C., Houshmandi, S. S., Olivas, W. M. & Bussemaker, H. J. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 17675–17680.
22. Foat, B. C., Morozov, A. V. & Bussemaker, H. J. (2006) *Bioinformatics*, in press.
23. van Steensel, B. & Henikoff, S. (2003) *BioTechniques* **35,** 346–350, 352–354, 356–357.
24. Negre, N., Hennetin, J., Sun, L. V., Lavrov, S., Bellis, M., White, K. P. & Cavalli, G. (2006) *PLoS Biol.* **4,** E170.
25. Smothers, J. F. & Henikoff, S. (2001) *Mol. Cell. Biol.* **21,** 2555–2569.
26. Kohonen, T. (2001) *Self-Organizing Maps* (Springer, Berlin).
27. MacAlpine, D. M., Rodriguez, H. K. & Bell, S. P. (2004) *Genes Dev.* **18,** 3094–3105.
28. Mito, Y., Henikoff, J. G. & Henikoff, S. (2005) *Nat. Genet.* **37,** 1090–1097.
29. Wang, J., Delabie, J., Aasheim, H., Smeland, E. & Myklebost, O. (2002) *BMC Bioinformatics* **3,** 36.
30. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V., *et al.* (2003) *Nucleic Acids Res.* **31,** 374–378.
31. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. (2004) *Nucleic Acids Res.* **32,** D91–D94.
32. Hanes, S. D. & Brent, R. (1989) *Cell* **57,** 1275–1283.
33. Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. & Desplan, C. (1989) *Cell* **59,** 553–562.
34. Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M. F., Rifkin, S. A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P. E., *et al.* (2004) *Science* **306,** 655–660.
35. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287,** 2185–2195.
36. Phair, R. D., Scaffidi, P., Elbi, C., Vecerova, J., Dey, A., Ozato, K., Brown, D. T., Hager, G., Bustin, M. & Misteli, T. (2004) *Mol. Cell. Biol.* **24,** 6393–6402.
37. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. (2002) *Mol. Cell* **10,** 1453–1465.
38. Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M. & Hoffman, A. R. (2006) *Science* **312,** 269–272.
39. Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E., *et al.* (2002) *Genome Biol.* **3,** RESEARCH0083.